



UNIVERSIDAD
COMPLUTENSE
MADRID

Proyecto de Innovación y Mejora de la Calidad Docente

Convocatoria 2015

Nº de proyecto 133

Título del proyecto:

Integración de software estadístico R en el entorno Moodle. Elaboración de casos prácticos de inferencia estadística aplicados a la Empresa y Comercio

Nombre del responsable del proyecto:

Francisco Javier Martín Campo

Centro:

Facultad de Ciencias Económicas y Empresariales

Departamento:

Estadística e Investigación Operativa II (Métodos de Decisión)

Contenido

Objetivos propuestos en la presentación del proyecto3

Objetivos alcanzados4

Metodología empleada en el proyecto.....5

Recursos humanos.....6

Desarrollo de las actividades.....7

Anexo8

Objetivos propuestos en la presentación del proyecto

En la solicitud del proyecto se plantearon los siguientes objetivos, clasificados en generales y específicos.

Objetivos generales:

1. Fomentar el aprendizaje autónomo del alumno en materias de inferencia estadística.
2. Fomentar la metodología e-learning mediante casos prácticos autoevaluables tras la finalización de su resolución.
3. Fomentar un ritmo de trabajo regular basado en la realización de ejercicios, con la finalidad de asimilar los contenidos de forma progresiva durante el desarrollo de la materia.
4. Elaboración de contenidos prácticos con enfoque real aplicado a la Empresa y Comercio atractivos.
5. Integración de los casos prácticos en la plataforma de Moodle para aprovechar las herramientas de autoevaluación.

Objetivos específicos:

1. Promover la participación del alumno en la plataforma Moodle.
2. Promover el autoaprendizaje del alumno mediante contenidos motivadores.
3. Promover el trabajo colectivo de los alumnos mediante casos prácticos más elaborados que exijan la participación colectiva del alumnado.
4. Incrementar la motivación del alumno en nuestras asignaturas.
5. Implementar los casos prácticos usando software estadístico R.
6. Incrementar la formación del alumno en manejo de otro tipo de software aplicado a resolver los casos prácticos elaborados.

Objetivos alcanzados

De los objetivos planteados, se ha alcanzado el objetivo principal: la implementación de casos prácticos usando software estadístico R adaptados a la plataforma Moodle. Su implementación ha llevado la mayor parte del proyecto, pero con ello se ha conseguido una batería de casos prácticos listos para su uso por parte de los alumnos en la plataforma Moodle, perfectamente integrados con las herramientas de calificación a disposición de los docentes.

La implementación ha sido el cuello de botella del proyecto. Una vez implementados casos vacíos (sin enunciado aparente), éstos sirven de plantilla para incorporar nuevos enunciados y, por tanto, una mayor base de problemas.

Todos aquéllos objetivos relacionados con el alumno no han podido ser alcanzados debido a que las asignaturas con contenidos de inferencia estadística se cursan en el segundo cuatrimestre. A partir de este momento, se desea tener una segunda fase del proyecto en el que se pueda poner a disposición del alumno la batería de casos prácticos creada con el fin de evaluar su utilidad y ventajas/inconvenientes de su uso.

Metodología empleada en el proyecto

Para conseguir los resultados de este proyecto, se han implementado vía software estadístico R varios casos prácticos enfocados a asignaturas de Empresa y Comercio, centrados en inferencia estadística, concretamente, casos prácticos para intervalos de confianza y contrastes de hipótesis.

Los casos prácticos han sido creados por profesores que imparten asignaturas con contenido de inferencia estadística en los Grados de Administración y Dirección de Empresas y Comercio.

La elaboración de casos prácticos ha tenido dos líneas de actuación:

1. Implementación informática de los casos usando software estadístico R.
2. Elaboración de una descripción motivadora e integrada en casos de posible uso real.

El trabajo conjunto de los casos prácticos en las dos líneas de actuación anteriormente presentadas desemboca en la integración como actividad evaluable dentro de la plataforma Moodle, con el principal objetivo de explotar al máximo las capacidades de la misma en cuanto a calificación se refiere.

El desarrollo del trabajo se ha estructurado en los siguientes niveles:

1. Definición de los casos prácticos a incluir.
2. Apertura de un seminario de trabajo con acceso a todos los profesores del equipo, para dar soporte a los casos prácticos.
3. Elaboración de casos prácticos enfocados a intervalos de confianza.
4. Elaboración de casos prácticos enfocados a contrastes de hipótesis de significación.
5. Descripción formal y rigurosa con fondo real de los casos creados.
6. Integración de los casos prácticos en Moodle.

Todo ello se ha elaborado en un espacio de trabajo dentro de la plataforma Moodle titulado “Proyecto PIMCD 133”, cuyo acceso es restringido a los autores del proyecto, pero previa petición, puede ser explorado por cualquier interesado/a. En el Anexo puede encontrarse un caso de estudio detallado.

Recursos humanos

Los recursos humanos con los que ha contado el proyecto han sido los profesores integrantes del mismo. Es importante señalar que el equipo se ha visto afectado por las bajas de 3 profesores, dos de ellas por finalización de contrato de modo imprevisto y otra por cambio de Universidad.

Por otro lado, el responsable del proyecto ha cambiado de centro dentro de la Universidad Complutense de Madrid, provocando también desajustes en los planes iniciales.

Aun así, la voluntad del equipo ha sido avanzar en los objetivos planteados aceptando la carga de trabajo extra que todos estos cambios han producido.

Desarrollo de las actividades

Para la consecución de los objetivos descritos en la propuesta, se han planteado las siguientes actividades:

1. Estudio del artículo de referencia en nuestro proyecto: "*Flexible generation of e-learning exams in R: Moodle quizzes, OLAT assessments, and beyond*", de A. Zeileis, N. Umlauf y F. Leisch en *Journal of Statistical Software* 58(1), pp.: 1-36 y creación de los primeros casos prácticos.
2. Reunión con el Profesor Antonio Alonso Ayuso de la Universidad Rey Juan Carlos que ya había trabajado antes en la elaboración de casos de contenido similar.
3. Implementación de casos prácticos relacionados con intervalos de confianza.
4. Validación de los casos prácticos implementados en intervalos de confianza.
5. Implementación de casos prácticos relacionados con contrastes de hipótesis de significación bilaterales y unilaterales.
6. Validación de los casos prácticos creados en contrastes de hipótesis de significación.
7. Incorporación de enunciados con contenido empresarial/económico a los casos implementados.

Anexo

A continuación se presentan capturas de pantalla de un caso práctico creado, con el fin de ilustrar el material creado por el equipo de trabajo.

El seminario virtual de trabajo presenta la siguiente estructura:

The screenshot displays the 'Proyecto PIMCD 133' interface. At the top, it shows 'Mi Campus ► seminario-invest-59704-4'. A left-hand navigation menu titled 'Navegación' lists various options, with 'seminario-invest-59704-4' expanded to show sub-items like 'Participantes', 'Insignias', 'Proyecto PIMCD 133', 'Ejemplos varios (Antonio Alonso)', 'Intervalos de Confianza (propios)', 'Contrastes de hipótesis bilaterales (propios)', 'Contrastes de hipótesis unilaterales (propios)', 'Intervalos de confianza con enunciado', 'Contrastes de hipótesis bilaterales con enunciado', and 'Contrastes de hipótesis unilaterales con enunciado'. The main content area is divided into three sections:

- Ejemplos varios (Antonio Alonso)**: A list of six items, each with a document icon and a checkmark:
 - Estadística Descriptiva: Boxplot + Regresión
 - Estadística descriptiva: Diagrama de dispersión
 - Estadística descriptiva: Modelo de regresión lineal simple
 - Inferencia Estadística: Intervalo de confianza
 - Inferencia Estadística: Intervalos y contrastes
- Intervalos de Confianza (propios)**: A list of nine items, each with a document icon and a checkmark:
 - Media (Varianza conocida)
 - Media (varianza desconocida)
 - Varianza (media desconocida)
 - Diferencia de medias (varianzas conocidas)
 - Diferencia de medias (varianzas desconocidas e iguales)
 - Razón de varianzas
 - Proporción
 - Diferencia de proporciones
- Contrastes de hipótesis bilaterales (propios)**: A list of two items, each with a document icon and a checkmark:
 - Media
 - Media Varianza Desconocida

Donde pueden apreciarse los distintos cuestionarios creados, correspondientes a los casos de estudio elaborados. Se cuenta con una serie de ejemplos creados por el Profesor Antonio Alonso Ayuso (URJC) que sirvieron de ejemplo para la creación de los casos basados en inferencia estadística.

A continuación hay tres bloques de cuestionarios: Intervalos de confianza, contrastes de hipótesis bilaterales y contrastes de hipótesis unilaterales con 8 casos diferentes cada uno de ellos. Cada cuestionario creado puede constar del número de problemas deseado, donde las diferencias están en los datos y no en la descripción del problema. De este modo, el profesor garantiza que todos los alumnos resolverán el mismo

problema pero con datos distintos, garantizando la igualdad entre los problemas asignados a los alumnos. Estos tres bloques están formados por problemas “huecos” a modo de plantilla, es decir, contienen un enunciado meramente formal.

Finalmente, se dispone de otros tres bloques con la misma estructura anterior cuyos problemas cuentan con enunciados con enfoque económico/empresarial.

A modo de ejemplo, se pretende ilustrar el funcionamiento de los mismos con capturas de pantalla. Primeramente el alumno accede a un cuestionario y es remitido al enunciado del mismo:

Pregunta 1

Sin responder aún

Puntúa como 10,00

▼ Marcar pregunta

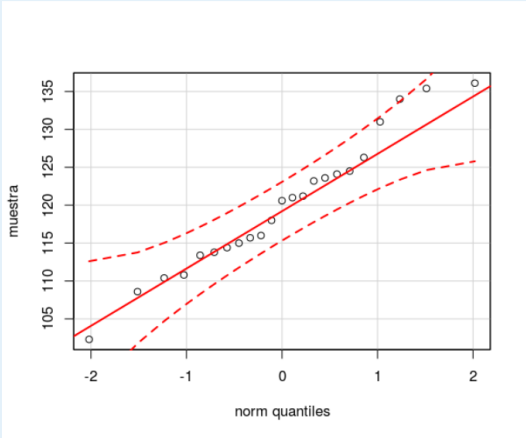
✎ Editar pregunta

Una agencia de alquiler de automóviles desea estimar el número medio de kilómetros diarios que realiza su flota de automóviles. A tal fin, para varios días de la semana toma los recorridos de $n = 23$ vehículos de su flota. Se supone la hipótesis de normalidad con desviación típica conocida e igual a 9 de la variable aleatoria número de kilómetros recorridos por día.

Para estudiar la variable aleatoria, se toma una muestra aleatoria simple de tamaño $n = 23$, $X = (x_1, \dots, x_n)$ de la que se sabe que:

$$\sum_{i=1}^{23} x_i = 2759.4 \quad \sum_{i=1}^{23} x_i^2 = 332769$$

Para asegurar que los datos proceden de una población normal, se ha estudiado el $Q-Q$ plot, un gráfico estadístico para estudiar la normalidad de los datos, obteniendo la siguiente representación:



Se puede observar que los valores de la muestra se encuentran dentro de las bandas de confianza, por lo que hay evidencia estadística para aceptar normalidad en los datos de la muestra.

De modo más riguroso, se realiza el contraste de Shapiro-Wilk para contrastar si la muestra se distribuye con normalidad, obteniendo un p -valor igual a 0.72, valor que siendo mayor

De modo más riguroso, se realiza el contraste de Shapiro-Wilk para contrastar si la muestra se distribuye con normalidad, obteniendo un p -valor igual a 0.72, valor que siendo mayor que el nivel de significación $\alpha = 0.01$ determina que hay evidencia estadística para aceptar que la muestra se distribuye bajo una distribución normal.

Se desea contrastar con un nivel de significación del 1% si el número de kilómetros diarios es distinto de 120.28. Para ello, se deben responder a las siguientes cuestiones:

1. De los siguientes estadísticos:

- | | |
|--|--|
| (a) $\frac{(\bar{x} - \mu_0) \sqrt{n}}{\sigma}$ | (b) $\frac{(\bar{x} - \mu_0) \sqrt{n}}{s_{n-1}}$ |
| (c) $\frac{n s_n^2}{\sigma_0^2}$ | (d) $\frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$ |
| (e) $\frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{\sqrt{\frac{n s_n^2}{n+m-2} + \frac{m s_m^2}{n+m-2}}}$ | (f) $\frac{s_n^2 - s_m^2}{s_{n+m-1}^2} \sigma_0^2$ |
| (g) $\frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$ | (h) $\frac{(p_1 - p_2) - \pi_0}{\sqrt{\frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{m}}}$ |

El estadístico de contraste es y su distribución de probabilidad

2. El contraste que se desea realizar es:

- H_0 :
- H_1 :

3. El valor del estadístico de contraste es:

4. De las siguientes opciones:

- (a) $R_c = \{|z_c| \geq |z_{\alpha/2}|\}$
- (b) $R_c = \{z_c \geq z_\alpha\}$
- (c) $R_c = \{z_c \leq z_\alpha\}$
- (d) $R_c = \{z_c \geq -z_\alpha\}$
- (e) $R_c = \{z_c \leq -z_\alpha\}$

la región crítica es la opción

5. En base a los resultados obtenidos, hay evidencia estadística para la hipótesis nula.

6. El p -valor de la muestra es:

Como se puede contemplar, el alumno tiene distintas opciones de respuesta. En este caso de estudio, menús desplegables e introducción de valores numéricos. Otros tipos de respuestas pueden ser incluidos de modo análogo (múltiple respuestas, etc.)

El alumno por tanto deberá responder a las preguntas planteadas:

El estadístico de contraste es **a** y su distribución de probabilidad **Normal Estándar**

2. El contraste que se desea realizar es:

$H_0: \mu = 120,28$

$H_1: \mu \neq 120,28$

3. El valor del estadístico de contraste es: **-0,17**

4. De las siguientes opciones:

(a) $R_c = \{|z_c| \geq |z_{\alpha/2}|\}$

(b) $R_c = \{z_c \geq z_{\alpha}\}$

(c) $R_c = \{z_c \leq z_{\alpha}\}$

(d) $R_c = \{z_c \geq -z_{\alpha}\}$

(e) $R_c = \{z_c \leq -z_{\alpha}\}$

la región crítica es la opción **a**

5. En base a los resultados obtenidos, hay evidencia estadística para **Rechazar** la hipótesis nula.

6. El p-valor de la muestra es: **0,05**

Proyecto PIMCD 2015 - 133

Siguiente

Una vez que el alumno termina el cuestionario, pulsando en el botón “siguiente” y aceptando el envío del cuestionario, el alumno puede comprobar la puntuación obtenida, sus respuestas y si éstas eran correctas o no.

Pregunta 1

Parcialmente correcta

Puntúa 8,00 sobre 10,00

⚑ Marcar pregunta

✎ Editar pregunta

Una agencia de alquiler de automóviles desea estimar el número medio de kilómetros diarios que realiza su flota de automóviles. A tal fin, para varios días de la semana toma los recorridos de $n = 23$ vehículos de su flota. Se supone la hipótesis de normalidad con desviación típica conocida e igual a 9 de la variable aleatoria número de kilómetros recorridos por día.

Para estudiar la variable aleatoria, se toma una muestra aleatoria simple de tamaño $n = 23$, $X = (x_1, \dots, x_n)$ de la que se sabe que:

$$\sum_{i=1}^{23} x_i = 2759,4 \quad \sum_{i=1}^{23} x_i^2 = 332769$$

Para asegurar que los datos proceden de una población normal, se ha estudiado el *Q-Q plot*, un gráfico estadístico para estudiar la normalidad de los datos, obteniendo la siguiente representación:

Se puede observar que los valores de la muestra se encuentran dentro de las bandas de confianza, por lo que hay evidencia estadística para aceptar normalidad en los datos de la muestra.

De modo más riguroso, se realiza el contraste de Shapiro-Wilk para contrastar si la muestra se distribuye con normalidad, obteniendo un p -valor igual a 0.72, valor que siendo mayor que el nivel de significación $\alpha = 0.01$ determina que hay evidencia estadística para aceptar que la muestra se distribuye bajo una distribución normal.

Se desea contrastar con un nivel de significación del 1% si el número de kilómetros diarios es distinto de 120.28. Para ello, se deben responder a las siguientes cuestiones:

1. De los siguientes estadísticos:

$$(a) \frac{(\bar{x} - \mu_0)\sqrt{n}}{\sigma} \quad (b) \frac{(\bar{x} - \mu_0)\sqrt{n}}{S_{n-1}}$$

$$(c) \frac{ns_1^2}{\sigma_0^2} \quad (d) \frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$$

$$(e) \frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{\sqrt{\frac{ns_1^2 + ms_2^2}{n+m}}} \quad (f) \frac{s_1^2 - s_2^2}{s_1^2 + s_2^2}$$

$$(g) \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} \quad (h) \frac{(p_1 - p_2) - \pi_0}{\sqrt{\frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{m}}}$$

El estadístico de contraste es y su distribución de probabilidad

2. El contraste que se desea realizar es:

$$H_0: \mu = 120.28$$

$$H_1: \mu \neq 120.28$$

3. El valor del estadístico de contraste es:

4. De las siguientes opciones:

$$(a) R_c = \{|z_c| \geq |z_{\alpha/2}|\}$$

$$(b) R_c = \{z_c \geq z_{\alpha}\}$$

$$(c) R_c = \{z_c \leq z_{\alpha}\}$$

$$(d) R_c = \{z_c \geq -z_{\alpha}\}$$

$$(e) R_c = \{z_c \leq -z_{\alpha}\}$$

la región crítica es la opción

5. En base a los resultados obtenidos, hay evidencia estadística para la hipótesis nula.

6. El p -valor de la muestra es:

Además, el alumno también puede ver la solución detallada como muestran las siguientes capturas de pantalla.

El contraste de hipótesis a realizar es el siguiente:

$$H_0: \mu = 120.28$$

$$H_1: \mu \neq 120.28$$

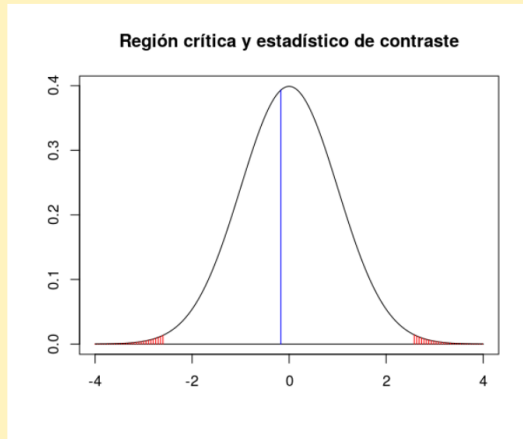
La media de la muestra dada es $\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{2759.4}{23} = 119.97$

Al tratarse de un contraste bilateral debemos tomar el valor $z_{\alpha/2} = 2.58$.

El estadístico de contraste viene dado por $z_c = \frac{(\bar{x} - \mu_0)\sqrt{n}}{\sigma} = \frac{(119.97 - 120.28)\sqrt{23}}{9} = -0.17$

La región crítica del contraste viene dada por $R_c = \{z_c \leq -z_{\alpha/2}\} \cup \{z_c \geq z_{\alpha/2}\} = \{|z_c| \geq |z_{\alpha/2}|\}$

Gráficamente se tiene:



donde la región marcada en rojo denota la región crítica y la línea azul denota el valor del estadístico de contraste.

Como NO se cumple la condición de la región crítica, existe evidencia estadística para ACEPTAR la hipótesis nula con el nivel de significación fijado.

El cálculo del p -valor del contraste es $P(|N(0,1)| > |z_c|) = 2P(N(0,1) > |z_c|) = 2P(N(0,1) > 0.17) = 2 \cdot 0.43251 = 0.86501$.

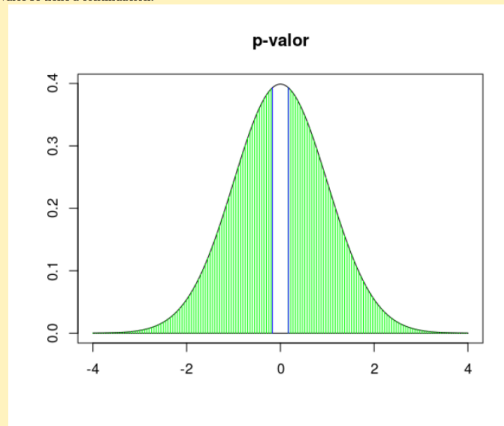
La interpretación gráfica del cálculo del p -valor se tiene a continuación:

donde la región marcada en rojo denota la región crítica y la línea azul denota el valor del estadístico de contraste.

Como NO se cumple la condición de la región crítica, existe evidencia estadística para ACEPTAR la hipótesis nula con el nivel de significación fijado.

El cálculo del p -valor del contraste es $P(|N(0,1)| > |z_c|) = 2P(N(0,1) > |z_c|) = 2P(N(0,1) > 0.17) = 2 \cdot 0.43251 = 0.86501$.

La interpretación gráfica del cálculo del p -valor se tiene a continuación:



Finalmente, se puede comprobar que el p -valor (0.86501) es MAYOR que el nivel de significación α (0.01) y, por tanto, existe evidencia estadística para ACEPTAR la hipótesis nula con el nivel de significación fijado.

Nota: El estudiante deberá interpretar la solución obtenida en el contexto definido en el enunciado.